# Context-Adaptive Sub-Nyquist Sampling for Low-Power Wearable Sensing Systems

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



# Context-Adaptive Sub-Nyquist Sampling for Low-Power Wearable Sensing Systems

Giovanni Schiboni*, Celia Martin Vicario*, Juan Carlos Suarez*, Federico Cruciani†, and Oliver Amft*

**Abstract**—This paper investigates a context-adaptive sample acquisition strategy at sub-Nyquist sampling rate for wearable embedded sensor devices. Our approach can be applied to compressive sensing frameworks to minimise sampling and transmission costs. We consider a context estimate to represent the local signal structure and a feed-forward response model to continuously tune signal acquisition of an online sampling and transmission system. To evaluate our approach, we analysed the performance in different pattern recognition scenarios. We report three case studies here: (1) eating monitoring based on electromyography measurements in smart eyeglasses, (2) human activity recognition based on waist-worn inertial sensor data, and (3) heartbeat detection and arrhythmia classification based on single-lead electrocardiogram readings. Compared to conventional sub-Nyquist sampling, our context-adaptive approach saves between 13% to 22% of energy, while achieving similar pattern recognition performance and reconstruction error.

**Index Terms**—Compressive sensing, physiological sensing, energy-efficient sensing, computation offloading, machine learning for healthcare.

## 1 Introduction

The embedded sensors, microcontroller ($\mu C$), and radio, of modern wearable devices enable users to monitor physiological and behavioural signals, which could support personalised healthcare [1]. However, the energy required to continuously operate wearable sensor devices is limiting their use. Off-loading computationally complex pattern recognition tasks to a sample processing unit, e.g. a smartphone or the computing cloud, does not reduce the energy needed for sampling and sample transmission.

Traditional sensor sampling implements a constant sampling frequency at or above the rate imposed by the renowned Shannon-Nyquist theorem. The constant sampling rate however implies an adequate energy budget that constrains device miniaturisation due to battery size, runtime between recharges, or even renders their application infeasible. Compressive Sensing (CS) [2] aims to reduce power and memory footprint for sensing applications by taking few random samples at the sensor, and by forwarding this sparse signal to a receiving unit that uncompresses the signal again for further processing [3]. While some variants of CS focus on data compression for storage and transmission [4], energy consumption in a sensor device could be reduced by controlling the rate at which the analog-to-digital converter operates [5]. The basic concept is to sample a signal at its information rate, rather than its maximum bandwidth, which is also called sub-Nyquist sampling (SNS) [6], [7]. SNS is a data acquisition strategy, where samples are collected uniformly, at random, below the Nyquist frequency. When uncompressing a SNS signal, the reconstruction quality

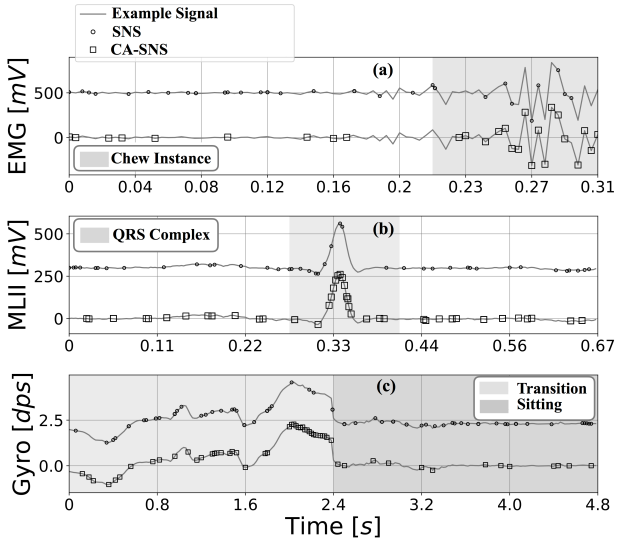


Fig. 1: Example sampling patterns of our proposed CA-SNS approach and conventional SNS: (a) EMG signal of the Temporalis muscle. (b) Single, modified lead II ECG of the heart. (c) Gyroscope signal at the waist during physical activity. Signal traces were offset to improve readability. Total sample counts are identical for SNS and CA-SNS in each subplot.


- *G. Schiboni, C. M. Vicario, J. C. Suarez, and O. Amft, are with the Chair of Digital Health, FAU Erlangen-Nürnberg, Germany. Email:{giovanni.schiboni, oliver.amft}@fau.de*
- *F. Cruciani is with the School of Computing, Ulster University, United Kingdom.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



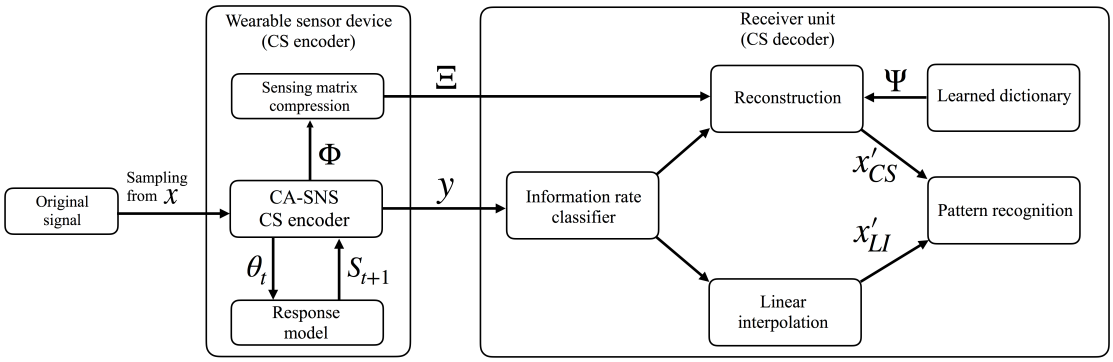


Fig. 2: Schematic of the proposed CA-SNS approach, including a wearable embedded sensor device (CS encoder) and a remote receiver unit (CS decoder), e.g. a smartphone or cloud computing, which can complement the resource-limited wearable device. The wearable sensor device derives time-dependent context estimate $\theta_t$ and controls SNS through the feed-forward response model $S_{t+1}$. Subsampled signal $\boldsymbol{y}$ and auxiliary vector $\boldsymbol{\Xi}$ are transmitted to the receiver unit. Depending on the outcome of an information rate classifier, the signal reconstruction is computed by a sparse reconstruction or a linear interpolation, which provide the approximated signals $x'_{\mathrm{CS}}$ and $x'_{\mathrm{LI}}$ respectively.

is related to the number of samples acquired, measured by the compression factor. The signal information rate of a source signal can be estimated by two families of functions connected by mathematical properties in any sparse domain [8], [9]: (1) Sparsity functions, related to compressibility, fairness, or dispersion measures; (2) Entropy functions, related to uncertainty, information, fuzziness, or complexity measures. Since signal information rate and signal compression factor are inversely related, compressibility will increase if information content in a signal decreases. With the constant uniform random sample distribution of conventional SNS, the quality of a reconstructed signal will vary with information rate.

As wearable sensor devices operate under varying context, their sensor signal information rate is rarely constant. With the constant uniform random sampling regime of SNS, the compression rate in a wearable sensor device varies, and so may the performance of downstream pattern recognition algorithms. To account for the varying context of a wearable device, compression settings must include safety margins. Therefore, current wearable sensor devices may acquire, store, and transmit many samples with minor information value. To maximise compression and minimise resource use, the number of sensor samples could be adjusted according to the dynamics of the signal.

In this paper, we introduce and analyse a context-adaptive sub-Nyquist sampling (CA-SNS) approach for wearable sensor devices that builds on CS to minimise energy consumption, by dynamically tuning the compression factor of the CS encoder. We illustrate how energy consumption can be reduced below the rates of conventional SNS while maintaining similar pattern recognition performance. In a context-adaptive sampling scenario, two main challenges can be identified. First, a context estimate is required to gauge the signal information rate. Second, a response model is needed to tune the compression factor based on the context estimate. The overall system behaviour policy must be implemented by the response model: Keep all energy-consuming components as long as possible in a low-power mode and only activate them if, according to the context estimate and some policy, relevant data must be sampled. Fig. 1 compares sampling patterns from different sensing modalities, where CA-SNS and SNS are applied. In CA-SNS, the temporal distribution of the sample sets is nonuniform, as the samples are clustered around regions of interest.

This paper provides the following contributions:

1) We derive a CA-SNS framework to obtain basic context estimates from sparsely sampled signals, and define a proportional feed-forward response model that approximates the signal information rate. While the response model continuously controls the CS stage, we propose a sensing matrix compression to inform the receiver unit.
2) We analyse and compare CA-SNS and SNS performances in three case studies: (1) eating event spotting based on electromyography (EMG); (2) human activity recognition based on accelerometer and gyroscope sampling; and (3) arrhythmia detection based on single-lead electrocardiography (ECG) sampling. We investigate the system performance tradeoff through simulations, considering pattern recognition, system energy, and memory footprint.

We formally introduce the CA-SNS sampling approach in Sec. 3 and relevant evaluation metrics in Sec. 4. Subsequently, low-power wearable sensor device case studies are analysed to explore the potential of CA-SNS in Secs. 5, 6, and 7. Lastly, we discuss CA-SNS results, opportunities, and current limitations in Sec. 8.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



## 2 Related Work

Adaptive CS strategies need a method for estimating a number of measurements required to reconstruct the original signal with acceptable quality. A common paradigm is to adjust the compression factor at the transmission stage based on an estimate of signal sparsity, and considering a bound on the signal reconstruction error. For example, Charalampidis et al. [10] introduced a framework for compressing an already acquired signal prior to transmission. The approach was based on a change point detection to adapt the compression ratio according to the time-varying sparsity level. The reconstruction error level was kept to a minimum to ensure quality of service requirements. Liu et al. [11] and He et al. [12] presented a compressive sensing framework to monitor ECG signals. An uniform sampling at full rate was adopted. Data were compressed before transmission, with a compression factor adapted according to the variation of the signal sparsity level. In order to define a relation between sparsity and optimal compression factor, a linear regression model was trained offline. Behravan et al. [13] proposed an adaptive CS framework to monitor biosignals. A sparsity variance monitoring mechanism was implemented in order to adapt the number of samples transmitted to the receiver. The adaptation was justified by the need to keep a low reconstruction error when the signal sparsity level varies. Tracking the signal information rate is more challenging, when the encoder does not have access to the original full signal, i.e., when CS would be applied for sensor sampling. By default, SNS samplers are unaware of signal sparsity. Thus, context estimates must be found that are representative to project ahead the further signal sparsity from an already sparsely acquired signal. A recent work by Traore et al. [14] introduced a CA-SNS scheme for cognitive radio applications. By employing the Bartlett method [15] as context estimate, the authors could choose sampling patterns according to spectrum changes. The sampling patterns were identified offline by a Sequential Forward Selection greedy-search algorithm. Fallahzadeh et al. [16] proposed a method to adaptively tune the sub-Nyquist rate at the sensing node according to the type of activity detected. As context estimate, a coarse-grained activity recognition was employed. A look-up table mapped the activity detected to a fixed sub-Nyquist rate, defined by an off-line heuristic optimisation tree. The rationale was to keep a fixed SNS rate until a change in activity was detected. In contrast to the attempts presented in the literature, our approach tunes the SNS rate in a sample-by-sample mode, thus adjusting the compression factor to the signal information rate and operating independently of predefined activity or context states.

## 3 Methods

The overall system architecture of our approach consists of a wearable embedded sensor device that samples and transmits compressed sensor data to a remote receiving unit. The receiver uncompresses the data and runs pattern recognition algorithms. Fig. 2 shows the architecture and information flow.

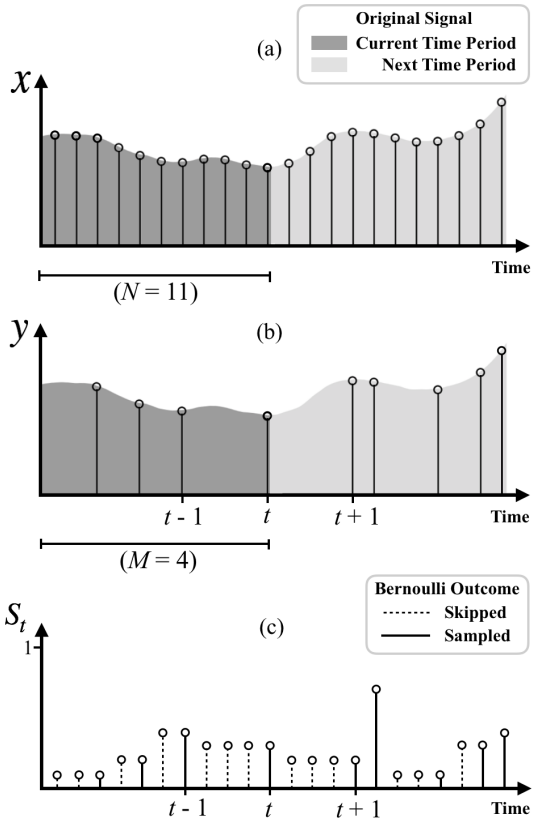


Fig. 3: (a) Full sampling mode. Equidistant sampling of the original signal $\boldsymbol{x}$. (b) CA-SNS sampling mode. Time points $(..., t\text{-}1, t, t\text{+}1, ...)$ represent the dynamic sampling interval and $\boldsymbol{y}$ the subsampled signal at the time points. Within the time period $N$, totally $M$ subsampled measurements are captured. (c) Sampling probabilities and Bernoulli outcome associated with each time sample in the CA-SNS approach.

The wearable device employs a sub-Nyquist sampler to acquire random samples at probability $S_t$ from the original signal represented as random variable $\boldsymbol{x}$. The context estimate $\theta_t$ approximates the signal information rate by extracting local characteristics of the signal. A feed-forward response model maps context estimate $\theta_t$ to the sampling probability $S_{t+1}$ to acquire a new sample. The subsampled signal $\boldsymbol{y}$ is transmitted with the auxiliary vector $\boldsymbol{\Xi}$. Signal information rate is classified at the receiver to bypass the CS decoder with a basic linear interpolation of $\boldsymbol{x}'_{\mathrm{LI}}$ when appropriate, i.e., information rate is low. For CS decoding, the approximated signal $\boldsymbol{x}'_{\mathrm{CS}}$ is derived from $\boldsymbol{y}$, $\boldsymbol{\Xi}$, and a learned overcomplete symbol dictionary $\boldsymbol{\Psi}$. The dictionary $\boldsymbol{\Psi}$ is derived based on training data. The approximated signals $\boldsymbol{x}'_{\mathrm{CS}}$ or $\boldsymbol{x}'_{\mathrm{LI}}$ are finally supplied to the pattern recognition stage.

Subsequently, we introduce the CA-SNS approach (Sec. 3.1), formalise the sparse signal measurement model according to CS principles (Sec. 3.2), introduce the reconstruction strategy (Sec. 3.3), and describe the dynamic sensing matrix compression method (Sec. 3.4).

### 3.1 CA-SNS Approach

In SNS, the sampling process is modelled according to a Bernoulli probability distribution, sampling the value 1, i.e., take a sample, with sampling probability $S$, and sampling the value 0, i.e., skip the sample, with probability $1 - S$. In CA-SNS, we consider a time-dependent sampling probability $S_t$, adjustable according to the signal information rate. Fig. 3 depicts an example sampling process.

We derived the sampling probability $S_t$ by computing a context estimate $\theta_t$, as a feature of the subsampled measurement signal $\boldsymbol{y}$, in particular:

$$\theta_t = f_C(y_t), \tag{1}$$

where $t$ is the current time-step, $y_t$ is the latest acquired sampled signal at $t$, $f_C(\cdot)$ implements the signal information rate approximation at the wearable sensor device.

A feed-forward proportional response model $f_\gamma(\cdot)$ converts $\theta_t$ to $S_{t+1}$:

$$S_{t+1} = f_\gamma(\theta_t) = S_l + \left[\frac{S_h - S_l}{\theta_h - \theta_l} \cdot \theta_t - \theta_l\right], \tag{2}$$

where $\theta_l$ and $\theta_h$ are lower and upper bounds of the context estimate interval. Sampling probability $S_t$ is set to the minimum $S_l$ if $\theta_t \leq \theta_l$. Similarly, the sampling probability $S_t$ is set to the maximum $S_h$ if $\theta_t \geq \theta_h$. The parameters $S_h$, $S_l$, $\theta_h$, and $\theta_l$ are tunable system design parameters and their choice affects the system sensitivity. Our CS encoder design builds on the assumption that the original signal $\boldsymbol{x}$ entails high signal information rate that is localised in some sampling intervals and that the intervals are identifiable from signal features, i.e., $\theta_t$. For many applications of wearable sensor devices we observed that elevated signal information rate is associated with increased signal energy. For example, in EMG signals, short bursts of high frequency components can be observed during charge and discharge of motor units during muscle work, however, signal frequency content and energy drops to baseline between individual contractions. In waist-worn inertial sensor data, sedentary phases with comparably low signal energy are interrupted by infrequent phases of activities of elevated intensity and signal energy. Even periodic patterns, such as QRS complexes of the ECG, are interspaced with low signal energy phases of variable duration. CA-SNS attempts to adjust sampling probability $S_t$ according to the signal information rate, and thus could deploy a dependent of signal energy to derive the context estimate $\theta_t$. Here, as context estimate we use $f_C(y_t) = var(y_t)$, where $var()$ is the incremental one-pass algorithm for computing sample variance [17], which updates the values of the mean and the sum of the squared deviations from that mean every time a new data sample is acquired. While the optimal number of samples needed to minimise the reconstruction error at the receiver unit is unknown, the feed-forward response model lets CA-SNS adapt the sampling strategy via $\theta_t$ and $S_{t+1}$ as a best guess from the present signal context.

### 3.2 Measurement Model

We consider the measurement model given by:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}, \tag{3}$$

where $\boldsymbol{y} \in \mathbb{R}^M$ is the signal vector composed of $M$ subsampled measurements, $\boldsymbol{x} \in \mathbb{R}^N$ is the original signal vector composed of $N$ samples (with $N >> M$), and $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ is the sensing matrix. The measurement model refers to a time period of length $N$, which represents the original signal vector $\boldsymbol{x}$ under quantisation limits, see Fig. 3. Note that during the time period $N$, $S_t$ may vary according to Eqs. 1, 2, and thus the sample interval varies too. The number of subsampled measurements $M$ is determined dynamically through $S_t$. We assume that the original signal $\boldsymbol{x}$ is sparse in some basis, i.e.,

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}, \tag{4}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{N \times G}$ is a matrix that represents the sparsifying basis. We interpret the sparsifying basis as an overcomplete dictionary, i.e., $N < G$. Eq. 4 describes a linear combination of the dictionary $\boldsymbol{\Psi}$ to represent the original signal $\boldsymbol{x}$, where $\boldsymbol{\alpha} \in \mathbb{R}^{G \times 1}$ is the column vector of weighting coefficients. The original signal $\boldsymbol{x}$ is called $K$-sparse when only $K$ elements of $\boldsymbol{\alpha}$ are non-zero, with $||\boldsymbol{x}||_0 \leq K << G$. Then, we can rewrite the measurement model as:

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha}, \tag{5}$$

with $\boldsymbol{\Phi}\boldsymbol{\Psi} \in \mathbb{R}^{M \times G}$ with $M << G$. Fig. 4 depicts the measurement model.

### 3.3 Reconstruction Strategy

A reconstructed version of the original signal can be computed as:

$$\boldsymbol{x}'_{CS} = \boldsymbol{\Psi}\boldsymbol{\alpha}', \tag{6}$$

where $\boldsymbol{\alpha}'$ is the recovered version of the weighting coefficient vector $\boldsymbol{\alpha}$.

Accurate reconstruction is possible only when $\boldsymbol{\Phi}$ satisfies the Uniform Uncertainty Principle, also known as Restricted Isometry Property (RIP) [18]. Since RIP is too difficult to validate directly, often a more easily computable requirement is used, based on the mutual coherence of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ given by:

$$\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \sqrt{N} \max_{1 \leq k,j \leq N} |\langle \phi_k, \psi_j \rangle|, \tag{7}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



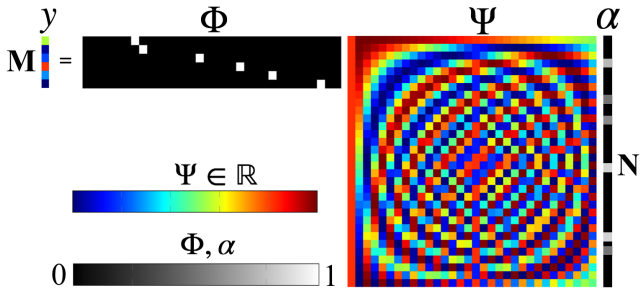


Fig. 4: Illustration of the CS measurement model. The sampled signal $\boldsymbol{y}$ is interpreted as the product of the sensing matrix $\boldsymbol{\Phi}$ and the original signal $\boldsymbol{x} = \boldsymbol{\Psi\alpha}$, where $\boldsymbol{\Psi}$ is a dictionary of elementary signals, and $\boldsymbol{\alpha}$ is a vector of sparse coefficients. The shaded squares in $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\alpha}$, illustrate different element amplitudes. For simplicity, an orthonormal dictionary $\boldsymbol{\Psi} \in \mathbb{R}^{N\times G}$ with $G = N$ is depicted, however in this work we use an overcomplete dictionary $\boldsymbol{\Psi}$, i.e. $N < G$.

where $1 < \mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) < \sqrt{N}$ is the range of coherence values between any row of $\boldsymbol{\Phi}$ and any column of $\boldsymbol{\Psi}$. According to Eq. 7, the mutual coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ is interpreted as the largest correlation between any element pair of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. In fact, the degree of signal compressibility can be determined as follows [19]:

$$m \geq C \cdot \mu^2(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \cdot K \cdot \log N, \tag{8}$$

where $C$ is some positive constant, and $m$ is the minimum number of measurements which provide an exact $\boldsymbol{\alpha}$ recovery with high probability. The target sparsity $\Omega$ refers to the number of non-zero entries in the weighting coefficient vector $\boldsymbol{\alpha}$.

Since our CS encoder has no access to the signal sparsity $K$, the empirical theorem expressed in Eq. 8 cannot be employed to derive a context estimate. The sampler's blindness towards the signal sparsity is the reason why we employed the feed-forward response model.

Our reconstruction strategy was composed by three elements: reconstruction algorithm, dictionary learning, and the information rate classifier.

#### 3.3.1 Reconstruction Algorithms

We considered three types of reconstruction algorithms:

a) Orthogonal Matching Pursuit (OMP) [20]:

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{\alpha}||_1, \tag{9}$$

$$\text{s.t. } ||\boldsymbol{\Phi\Psi\alpha} - \boldsymbol{y}||_2 = 0. \tag{10}$$

b) Gradient projection-based sparse rec. (GPSR) [21]:

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{\Psi\Phi\alpha} - \boldsymbol{y}||_2^2 + \tau||\boldsymbol{W\alpha}||_1, \tag{11}$$

where $\tau \in \mathbb{R}_{>0}$ represents the relative weight of the $\ell_1$-norm and $\ell_2$-norm in the cost function, and $\boldsymbol{W}$ is a diagonal matrix composed by the free parameters $\omega_1, \ldots, \omega_n$ on the diagonal.

c) Iterative Shrinkage-Thresholding Alg. (ISTA) [22]:

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{\Psi\Phi\alpha} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{\alpha}||_1. \tag{12}$$

When a low signal information rate was detected at the receiver, we reconstructed the signal using linear interpolation.

#### 3.3.2 Dictionary Learning

Overcomplete dictionaries, created with data-driven techniques, have been explored previously, see for example [23]. Considering a set of $z$ training signals $\boldsymbol{x}_z = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_z] \in \mathbb{R}^{N\times z}$, the aim of the training is to find an optimal dictionary $\boldsymbol{\Psi} \in \mathbb{R}^{N\times G}$, which achieves sparse representations of the training signals in a set of coefficients $\boldsymbol{\beta} \in \mathbb{R}^{G\times z}$. Here, dictionaries were created using the K-SVD algorithm introduced by Aharon et al. [24].

#### 3.3.3 Information Rate Classifier

A low information rate of the original signal, a low sensitivity of the feed-forward model, or a combination of both conditions, often trigger a featureless measurement signal $\boldsymbol{y}$ characterised by: (a) all $M$ samples close to the noise level; or (b) insufficient $M$ to reconstruct the signal i.e., $M << m$. With insufficient information, reconstruction error gets inacceptable [18], which results in arbitrary weighted linear signal combination from the dictionary $\boldsymbol{\Psi}$. To avoid arbitrary reconstruction, we chose to apply a linear interpolation for a featureless signal $\boldsymbol{y}$. We designed an information rate classifier $f_Q(\boldsymbol{y})$ for the receiver based on the heuristic rule:

$$f_Q(\boldsymbol{y}) = \begin{cases} x'_{CS}, & \text{if } (M > F_M) \vee \\ & (\max(\boldsymbol{y}) > F_y), \\ x'_{LI}, & \text{otherwise,} \end{cases} \tag{13}$$

where $F_M \in \mathbb{N}$ is a sample threshold and $F_y \in \mathbb{R}$ is a value threshold. The approximated signal $x'_{CS}$ is computed by the CS decoder and the approximated signal $\boldsymbol{x}'_{\mathrm{LI}}$ is computed by linear interpolation.

### 3.4 Sensing Matrix Compression

At sparsity $K$ = const, the lower the mutual coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$, the lower the number of measurements needed for accurate reconstruction. Following the concept shown in Fig. 4, incoherent measurements with the sparsifying basis $\boldsymbol{\Psi}$ result in an unstructured sensing matrix $\boldsymbol{\Phi}$. For example, independent and identically distributed random matrices show a low mutual coherence of $\sim(2\log_2 N)^{\frac{1}{2}}$ with a variety of fixed sparsifying bases [18]. In particular, random sampling provides a sensing matrix $\boldsymbol{\Phi}$ that produces low mutual coherence $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$.

In order to compute $\boldsymbol{x}'_{CS}$, the sensing matrix $\boldsymbol{\Phi}$, related to the current time period $N$, is needed by the receiver. Transmitting the entire matrix, or just the coordinates of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



the non-zero positions, would entail an expensive energy overhead at the transmission stage. Here, we describe a method to compress $\mathbf{\Phi}$, which can sufficiently reduce the data transmission load and as well as minimise computational and memory footprint at the wearable sensor device side. To transmit the information of $\mathbf{\Phi}$, we use ranking and unranking combinations and lexicographic ordering [25], [26]. Lexicographic ordering creates a bijective mapping between a set of combinations, interpreted as column vectors of $\mathbf{\Phi}$, and an auxiliary vector $\mathbf{\Xi}$ of corresponding ranks. The auxiliary vector $\mathbf{\Xi}$ exclusively identifies the sensing matrix $\mathbf{\Phi}$ and is transmitted to the receiver at the end of each time period $N$. At the receiver, the sensing matrix $\mathbf{\Phi}$ is reconstructed from $\mathbf{\Xi}$.

Our approach includes the following steps:

1) Consider a subsampled measurement signal $\boldsymbol{y}$ with length $M$. For random sampling of 1-D signals, the entries of the sensing matrix $\mathbf{\Phi}$ are all zeros except for $M$ entries in $M$ different positions, i.e. $\mathbf{\Phi}^{ind}_{m,n} \neq 0$, with rows $m$ and columns $n$. Rows in the sensing matrix $\mathbf{\Phi}$ are ordered, i.e., sample position column increases with the matrix row (see Fig. 4 for an illustration).

As an example, consider the matrix:

$$\mathbf{\Phi} = \tag{14}$$

$$\begin{bmatrix} 0 & 1_{1,2} & \dots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \dots & 1_{2,6} & \dots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \dots & 1_{3,15} & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 1_{4,42} & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 1_{5,61} & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 1_{6,71} & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1_{7,79} & 0 \end{bmatrix},$$

with $\mathbf{\Phi} \in \mathbb{R}^{7\times 80}$, where $M = 7, N = 80$.

2) Partition $\mathbf{\Phi}$ into $J$ matrices as:

$$\mathbf{\Phi} = \left[ \begin{array}{c|c|c|c|c} \mathbf{\Phi}_1 & \cdots & \mathbf{\Phi}_j & \cdots & \mathbf{\Phi}_J \end{array} \right], \tag{15}$$

with $\mathbf{\Phi}_j \in \mathbb{R}^{M\times N/J}$ and $j \in \mathbb{N}$.

Following the above example, consider the matrix:

$$\mathbf{\Phi} = \left[ \begin{array}{c|c} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 \end{array} \right], \tag{16}$$

with $\mathbf{\Phi}_j \in \mathbb{R}^{7\times 40}$, where $J = 2$.

3) Transform $\mathbf{\Phi}_j$ elements as:

$$\mathbf{\Phi}_{j,m,n} = \begin{cases} \Lambda(j,m,n), & \text{if } \mathbf{\Phi}_{j,m,n} = 1 \\ 0 & \text{otherwise,} \end{cases} \quad \forall\, j \leq J, \quad m \leq M, \quad n \leq N/J. \tag{17}$$

where

$$\Lambda(j,m,n) = n - (j-1)\frac{N}{J} + (m-1)N \tag{18}$$

converts non-zero $\mathbf{\Phi}$ element subscripts into partition matrix $\mathbf{\Phi}_j$ linear indices.

Following the above example, matrices $\mathbf{\Phi}_j \in \mathbb{R}^{7\times 40}$ are computed as:

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0 & 2 & \dots & 0 & 0 & 0 \\ 0 & \dots & 46 & \dots & 0 & 0 \\ 0 & 0 & \dots & 95 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{\Phi}_2 = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 122 & \dots & 0 & 0 & 0 \\ 0 & \dots & 181 & \dots & 0 & 0 \\ 0 & 0 & \dots & 231 & \dots & 0 \\ 0 & 0 & 0 & \dots & 279 & 0 \end{bmatrix}.$$

4) Serialise $\mathbf{\Phi}_j$ partition matrices as:

$$\mathbf{\Phi}^{\text{vec}}_j = vec(\mathbf{\Phi}_j) \quad \forall\, j \leq J, \tag{19}$$

where the vectorisation function $vec()$ transforms the bi-dimensional sensing partition matrices $\mathbf{\Phi}_j$ in a 1-D vector $\mathbf{\Phi}^{\text{vec}}_j$, by executing the operation $\mathbb{R}^{M\times N/J} \rightarrow \mathbb{R}^{MN/J}$. Following the above example, each matrix $\mathbf{\Phi}_j \in \mathbb{R}^{7\times 40}$ is transformed into a vector $\mathbf{\Phi}^{\text{vec}}_j \in \mathbb{R}^{280}$.

5) Obtain non-zero element positions from $\mathbf{\Phi}^{\text{vec}}_j$ as:

$$\mathbf{\Phi}^{\text{nz}}_j = \{\Phi^{\text{vec}}_{j,q} | \Phi^{\text{vec}}_{j,q} \neq 0\} \quad \begin{array}{l} \forall\, j \leq J, \\ \forall\, q \leq MN/J. \end{array} \tag{20}$$

Following the above example, vectors $\mathbf{\Phi}^{\text{vec}}_1$ and $\mathbf{\Phi}^{\text{vec}}_2$ result in:

$$\begin{aligned} \mathbf{\Phi}^{\text{nz}}_1 &= [2, 46, 95], \\ \mathbf{\Phi}^{\text{nz}}_2 &= [122, 181, 231, 279]. \end{aligned} \tag{21}$$

6) Compress the vectors by computing their ranks:

$$\Phi^{\text{rank}}_j = f_\delta(\mathbf{\Phi}^{\text{nz}}_j) = \sum_{i=0}^{|\mathbf{\Phi}^{nz}_j|-1} \sum_{v=\Phi^{nz}_{j,i-1}+1}^{\Phi^{nz}_{j,i}-1} \binom{|\mathbf{\Phi}^{ind}_j| - (v+1)}{|\mathbf{\Phi}^{nz}_j| - (i+1)}, \quad \forall\, j \leq J. \tag{22}$$

The ranking function $f_\delta(\cdot)$ encodes the index combination $\mathbf{\Phi}^{\text{nz}}_j$ into an integer value, interpreted as lexicographic order.

Following the above example, the two vectors $\mathbf{\Phi}^{\text{nz}}_1$ and $\mathbf{\Phi}^{\text{nz}}_2$ are ranked as:

$$\begin{aligned} \Phi^{\text{rank}}_1 &= 77650, \\ \Phi^{\text{rank}}_2 &= 226141970. \end{aligned} \tag{23}$$

Tab. 2 reports the number of arithmetical operations for $\mathbf{\Phi}^{nz}_1$ and $\mathbf{\Phi}^{nz}_2$ ranking, derived according to Eq. 22.

7) Transmit the vector:

$$\mathbf{\Xi} = [\Phi^{\text{rank}}_1, \cdots, \Phi^{\text{rank}}_J]. \tag{24}$$

In the example, the set of ranks to be transmitted is:

$$\boldsymbol{\Xi} = [77650, 226141970]. \tag{25}$$

The transmission of $\Phi_1^{\text{rank}}$ requires 16.2 bits, or 2.03 Bytes, with an equivalent memory demand for 0.51 samples at 32-bit resolution. The transmission of $\Phi_2^{\text{rank}}$ requires 27.8 bits, or 3.48 Bytes, equivalent to a memory demand of 0.87 samples at 32-bit resolution. The total transmission for the time period $N$ in this example is $\sim 8.4$ samples, which is the sum of the M=7 samples and the two ranks, resulting in a net overhead of $\sim 20\%$, and, given $N$, an overall compression of $\sim 89\%$.

More generally, the number of possible permutations with no repetition for each sensing partition matrix $\boldsymbol{\Phi}_j$ is:

$$U_j = \frac{|\boldsymbol{\Phi}_j|!}{(|\boldsymbol{\Phi}_j| - |\boldsymbol{\Phi}_j^{nz}|)!}, \tag{26}$$

where $|\boldsymbol{\Phi}_j| = M \cdot N/J$ represents the number of partition matrix elements and $|\boldsymbol{\Phi}_j^{nz}|$ is the number of non-zero sensing partition matrix elements. When $|\boldsymbol{\Phi}_j^{nz}| = N/J$, $U_j$ represents the upper bound on the number of possible permutations with no repetition for a sensing partition matrix $\boldsymbol{\Phi}_j$. The memory demand $U^b$ to represent $\boldsymbol{\Xi}$ is:

$$U^b = \sum_j U_j^b \text{ with} \tag{27}$$

$$U_j^b = \log_2 U_j, \tag{28}$$

where $U_j^b$ is the number of bits to represent $\Phi_j^{rank}$.

In the Appendix, memory demand $U^b$ and energy consumption $E_{f_\delta}^{\mu C}$ related to the sensing matrix compression are further analysed.

8) To compute the original sensing matrix at the receiver, the inverse procedure must be applied. Firstly, unrank the $\boldsymbol{\Xi}$ elements to the correspondent vectors $\boldsymbol{\Phi}_j^{\text{nz}}$. Secondly, compute each sensing partition matrix $\boldsymbol{\Phi}_j$ from their corresponding vector $\boldsymbol{\Phi}_j^{\text{vec}}$. Finally, compute the original sensing matrix $\boldsymbol{\Phi}$ by concatenating the sensing partition matrices $\boldsymbol{\Phi}_j$, with $j = 1, \ldots, J$. Although various lexicographic ordering approaches may be adopted, in this paper we followed the ranking and unranking procedure proposed by Er [27].

## 4 Evaluation Methods

### 4.1 Simulation

To evaluate the dynamic performance of CA-SNS, we performed simulations with actual sensor data and approximated hardware characteristics. We adopted the finite-state simulation approach used by Buschhoff et al. [28], to reproduce dependencies among different components and between hardware and software.

| Number of machine cycles per operation | | | | | |
|---|---|---|---|---|---|
| $n_{\text{Add}}^{cyc}$ | $n_{\text{Mult}}^{cyc}$ | $n_{\text{Div}}^{cyc}$ | $n_{\text{Root}}^{cyc}$ | $n_{\text{Comp}}^{cyc}$ | $n_{\text{Exp}}^{cyc}$ |
| 60 | 50 | 80 | 380 | 12 | 210 |

TABLE 1: Number of machine cycles $n_o^{cyc}$ for the arithmetical operation $o$ on the simulated wearable device $\mu C$, i.e., ARM CortexM3.

### 4.2 Recognition Performance

In the case studies, we investigate two typical time series pattern recognition tasks: pattern spotting and pattern detection by continuous classification. To assess the pattern spotting performance, we consider standard retrieval metrics Precision and Recall. For the classification analysis, we compute balanced accuracy to minimise bias on class-imbalanced data as average of the ratio of correct over the total instances per class.

### 4.3 Execution Time

We quantify computational complexity to estimate a function execution time, where a function comprises the arithmetic operations, additions and subtractions (**Add**), multiplications (**Mult**), divisions (**Div**), square roots (**Root**), exponentials (**Exp**), and comparisons (**Comp**). Tab. 1 lists the number of machine cycles, for each type of arithmetical operation, related to the typical $\mu C$ considered in this work, i.e., ARM CortexM3. Moreover, Tab. 2 presents a breakdown of the functions $f_C, f_\gamma$, and $f_\delta$, executed at the sensor node.

The number of machine cycles $n_f^{cyc}$ to execute a function is then:

$$\begin{aligned} n_f^{cyc} =& (n_{f,\textbf{Add}}^{op} \cdot n_{\textbf{Add}}^{cyc}) + (n_{f,\textbf{Mult}}^{op} \cdot n_{\textbf{Mult}}^{cyc}) + \\ & (n_{f,\textbf{Div}}^{op} \cdot n_{\textbf{Div}}^{cyc}) + (n_{f,\textbf{Root}}^{op} \cdot n_{\textbf{Root}}^{cyc}) + \\ & (n_{f,\textbf{Exp}}^{op} \cdot n_{\textbf{Exp}}^{cyc}) + (n_{f,\textbf{Comp}}^{op} \cdot n_{\textbf{Comp}}^{cyc}), \end{aligned} \tag{29}$$

where $f \in [f_C, f_\gamma, f_\delta]$, and $n_{f,o}^{op}$ is the number of executions of an arithmetical operation $o$.

A function execution time $ET_f$ was derived by normalising the number of machine cycles $n_f^{cyc}$ by the $\mu$C's clock frequency $\nu$:

$$ET_f = \frac{n_f^{cyc}}{\nu}. \tag{30}$$

### 4.4 $\mu C$ Energy Consumption

Typical embedded $\mu C$s support active and standby states. The average energy consumption $EC_f^{\mu C}$ of a $\mu C$ to execute function $f$ was modelled as:

$$EC_f^{\mu C} = ET_f \cdot I_{\text{act}}^{\mu C} \cdot V^{\mu C}, \tag{31}$$

where $I_{\text{act}}^{\mu C}$ is the $\mu C$ average current consumption in active state, and $V^{\mu C}$ is the $\mu C$ voltage. The average energy consumption of a $\mu C$ in stand-by state was modelled as:

$$EC_{\text{stb}}^{\mu C} = T_{\text{stb}}^{\mu C} \cdot I_{\text{stb}}^{\mu C} \cdot V^{\mu C}, \tag{32}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



| Algorithm | Function ($f$) | N. Arithmetical Operations | | | |
|---|---|---|---|---|---|
| | | Add | Mult | Div | Comp |
| Adaptive sampling | $f_C$ (Context estimate) | $5gM$ | $2gM$ | $gM$ | $gM$ |
| | $f_\gamma$ (Response output) | $4M$ | $M$ | $M$ | $M$ |
| Sensing Matrix Compression | $f_\delta$ (Ranking) | For the example in Sec. 3.4 | | | |
| | | 2880 | 799 | 799 | 2806 |

TABLE 2: Breakdown of operations executed at the sensor node within a time period $N$. Variable $g$ denotes the number of sensor channels and $M$ the number of taken samples. For matrix ranking, the values shown here refer to the example in Sec. 3.4. The number of arithmetical operations for the sensing matrix compression, i.e., computing the ranking function of Eq. 22, were derived from counters throughout all simulations.

| Functionality | Component | Electric characteristics | | | | |
|---|---|---|---|---|---|---|
| | | $I_{\text{act}}$[mA] | $I_{\text{stb}}$[mA] | V | $\nu$ [Hz] | $\rho$ [bits] |
| $\mu$C | ARM CortexM3 [29] | 7.0 | 0.55 | 3.3 | $48 \cdot 10^6$ | 32 |
| Sensor | EMG (Actiwave, CamN-tech [30]) | 4.0 | 0.008 | 3.3 | 256 | 32 |
| | Accelerometer (MPU-9250 Invensens [31]) | 0.45 | 0.008 | 2.5 | 50 | 32 |
| | Gyroscope (MPU-9250 Invensens [31]) | 3.2 | 0.008 | 2.5 | 50 | 32 |
| | ECG (TI ECG Data Acquisition System [32]) | 0.2 | 0.006 | 3 | 360 | 32 |
| Radio Transmission | Radio transceiver TX= 0 dB m for 114 Bytes transfer payload (Chipcon CC2420 [33]) | 21.7 | 0.005 | 3 | - | - |

TABLE 3: Example electrical component characteristics to estimate energy consumption in our simulations. $I_{\text{act}}$: current consumption in active mode. $I_{\text{stb}}$: current consumption in stand-by mode. V: voltage. $\nu$: sampling frequency. $\rho$: resolution. For simplicity, a 32-bit sample resolution was considered for all sensors.

where $T_{\text{stb}}^{\mu C}$ is the $\mu C$ inactive time and $I_{\text{stb}}^{\mu C}$ is the $\mu C$ average current consumption in stand-by state.
The average total energy consumption of a $\mu C$ was modelled as:

$$EC^{\mu C} = EC_{\text{stb}}^{\mu C} + \sum_{tp} \sum_{f} EC_f^{\mu C}, \tag{33}$$

where $tp$ is the number of times a time period $N$ was processed.

### 4.5 Sensor Energy Consumption

The average instantaneous energy consumption $EC_t^s$ of a sensor operation was derived by:

$$\begin{aligned} I_t^s &= I_{\text{act}}^s \cdot D_t + I_{\text{stb}}^s \cdot (1 - D_t), \\ EC_t^s &= I_t^s \cdot V^s \cdot t_r, \end{aligned} \tag{34}$$

where $I_{\text{act}}^s$ and $I_{\text{stb}}^s$ are the average sensor current consumption in active and standby states respectively, $D_t$ is the instantaneous duty-cycle value, $V^s$ is the sensor voltage, and $t_r$ is the temporal resolution expressed in hours. The average total energy consumption of a sensor operation was modelled as:

$$EC^s = \sum_t EC_t^s. \tag{35}$$

### 4.6 Radio Transmission Energy Consumption

To estimate the energy consumption of the wireless communication, we relied on the energy model described by Prayati et al. [34]. The model considered the following three stages for transmission: (1) Initialisation of transmission and transferring data from memory to the radio chip FIFO buffer; (2) Back-off timeout; (3) Packet $p$ transmission via the wireless channel. Packet $p$ transmission energy $EC_p^r$ is:

$$EC_p^r = I_{\text{act}}^r \cdot V^r \cdot T_{\text{trans}}, \tag{36}$$

where $T_{\text{trans}}$ denotes the time to prepare and transmit a given data packet size, and $V^r$ is the radio transceiver

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



voltage. The average energy consumption of the radio transmitter in stand-by mode was modelled as:

$$EC^r_{\text{stb}} = T^r_{\text{stb}} \cdot I^r_{\text{stb}} \cdot V^r, \tag{37}$$

where $T^r_{\text{stb}}$ is the radio transmitter inactive time and $I^r_{\text{stb}}$ is the average current in stand-by mode. The average total energy consumption of radio transmission was modelled as:

$$EC^r = EC^r_{\text{stb}} + \sum_p EC^r_p. \tag{38}$$

### 4.7 Total Energy Consumption

The total energy consumption of the wearable device was derived by:

$$EC = EC^{\mu C} + EC^s + EC^r. \tag{39}$$

The energy saving was calculated as the difference between the total energy consumption EC in full sampling mode and the total energy consumption in sub-Nyquist sampling mode, i.e., CA-SNS or SNS.

### 4.8 Memory Demand

We derived memory demand $U^s$ for the acquired samples as:

$$U^s = \frac{\rho}{8} \cdot \sum_{tp} M, \tag{40}$$

where $\rho$ is the sample resolution. The total memory demand for data to be transmitted was derived as:

$$MD = U^s + U^b. \tag{41}$$

### 4.9 Reconstruction Quality

We employed a standard metric used in sparse reconstruction evaluation, i.e., Percentage Root-mean-square Difference (PRD), to quantify the reconstruction performance:

$$PRD = 100 \cdot \sqrt{\frac{\sum_{i=1}^{n} (\boldsymbol{x}(i) - \boldsymbol{x'}(i))^2}{\sum_{i=1}^{n} \boldsymbol{x}^2(i)}}, \tag{42}$$

where $\boldsymbol{x}'$ is the reconstructed signal.

### 4.10 Evaluation Procedure

We compared SNS performance with $S$ =const, and CA-SNS, by tuning the response model sensitivity. For each case study, we show results related to: (1) The tradeoff between recognition performance and energy saving, when sampling modalities are applied; (2) Absolute values of energy consumption EC and memory demand MD; (3) PRD of the reconstructed signals. Details of applied validation strategies are described in each case study section. Table 3 lists the component characteristics used in the case study simulations.

Due to the random nature of CA-SNS and SNS, the evaluation on each validation set was repeated ten times and results averaged. Statistical significance of the performance differences between CA-SNS and SNS was analysed with the corrected paired Student's test [35] at $5\%$ significance level.

## 5 Case Study: Spotting Chewing Patterns Using EMG-Monitoring Eyeglasses

3D-printed smart eyeglasses with EMG electrodes bilaterally integrated at the temple ear bends [36] were devised to monitor food chewing in a free-living study [37]. As reference for EMG sensing, we consider a standard 2-channel EMG recorder (Actiwave, CamNtech [30]).

### 5.1 Pattern Recognition Method

To spot eating events, Zhang and Amft [38] extracted features with non-overlapping 13 $s$ sliding windows from the continuous EMG measurements and used a one-class Support Vector Machine (oc-SVM) for event identification. Details of the method can be found in [38].

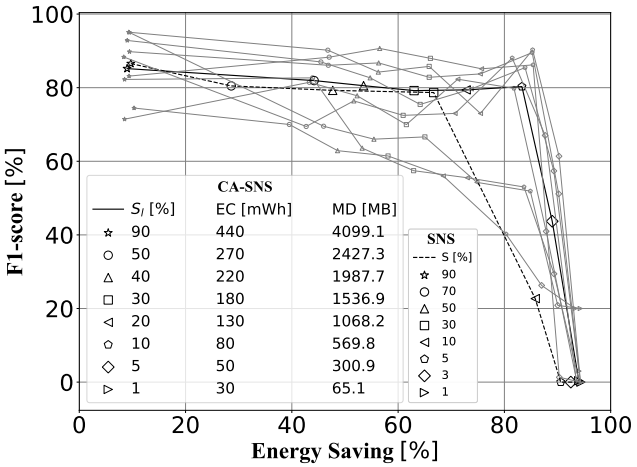


Fig. 5: Case study of Sec. 5, showing chewing pattern spotting results derived on the validation set across participants. Grey lines correspond to performance on indiviidual participant data. In SNS, $S$ was varied. In CA-SNS, $S_l$ was varied. Parameters: $N = 128$, $G = 6400$, $S_h = 1$, $\theta_h = 1000$, $\theta_l = 100$, $J = 10$, $F_M = 20$, $F_y = 75$, $\Omega = 25$.

### 5.2 Sensor Dataset

Ten healthy volunteers (4 females, 6 males) aged between 20 and 30 years wore the EMG-monitoring eyeglasses for one day. The eyeglasses were attached after getting up in the morning and kept on till bed time. Participants manually logged the occurrence of eating events in a diet journal with a one minute resolution. EMG data was collected at uniform sampling rate, i.e., 256 Hz per channel. A method was proposed to obtain the annotated eating events by combining reference EMG data and the participant journal [37].

### 5.3 Validation Strategy

We performed sample-wise evaluation of the pattern spotting performance against annotated eating events.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



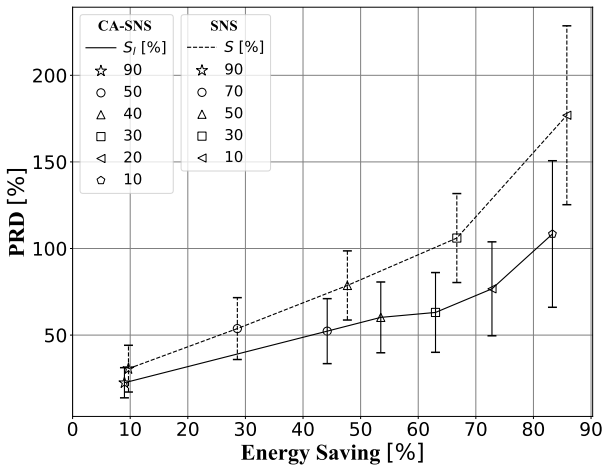


Fig. 6: Case study of Sec. 5. Parameters: $N = 128$, $G = 6400$, $S_h = 1$, $\theta_h = 1000$, $\theta_l = 100$, $J = 10$, $F_M = 20$, $F_y = 75$, $\Omega = 25$. Results derived on the validation set.

A leave-one-participant-out cross-validation was applied on data collected from nine participants, i.e., eight participants for training set, and one participant for validation set. The ocSVM's hyperparameter optimisation was performed using grid search. The best hyperparameter combination was applied to the tenth, so far left out participant, and final results averaged over all cross-validation folds. The CA-SNS hyperparameters were set manually, constant for all validation folds.

| | **Full sampling** mean (std) | **SNS** mean (std) | **CA-SNS** mean (std) | **Stat. Sig. Test** (Significance Level = 5%) |
|---|---|---|---|---|
| F1-Score [%] | 85.1 (1.3) | 79.3 (5.8) | 81.4 (7.5) | Not significant |
| PRD [%] | - | 108.6 (14.3) | 108.1 (8.4) | Not significant |
| Energy Saving [%] | - | 65.4 (0.1) | 82.3 (1.6) | $10^{-5}$ |
| EC [mWh] | 478.6 (10.2) | 95.6 (1.1) | 78 (2.4) | $10^{-5}$ |
| MD [MB] | 4524.4 (110.1) | 680.2 (1.2) | 558.1 (2.3) | $10^{-6}$ |

TABLE 4: Case study of Sec. 5, summarising chewing pattern spotting results derived on testing data. Results were averaged across ten evaluations, reporting mean and standard deviation (std). Parameters: $N = 128$, $G = 6400$, $S_h = 1$, $S_l = 10$, $S = 30$, $\theta_h = 1000$, $\theta_l = 100$, $J = 10$, $F_M = 20$, $F_y = 75$, $\Omega = 25$.

### 5.4 Results

The three reconstruction algorithms showed similar performance, see Appendix Fig. 3. We focused on results based on OMP. Fig. 5 compares CA-SNS and SNS for the retrieval performance and energy saving tradeoff. The CA-SNS clearly outperformed SNS, as the SNS F1-score dramatically dropped above an energy saving of 65%. At an energy saving of $\sim 80\%$, the F1-score difference was approx. $40\%$. Fig. 6 compares the PRD tradeoff between CA-SNS and SNS. The CA-SNS achieved a lower PRD than SNS for any minimum sampling probability $S_l$. Tab. 4 shows the results summary derived on the test dataset. At similar retrieval performance, CA-SNS saved approx. $17\%$ in energy and memory compared to SNS.

## 6 Case Study: Human Activity Recognition Using Waist-Worn Inertial Sensors

We analyse pattern classification based on physical activities from waist-worn inertial sensor data, i.e., accelerometer and gyroscope signals.

### 6.1 Pattern Recognition Method

We employed a 3-layer Convolutional Neural Network architecture used for activity classification by Cruciani et al. [39]. Segmentation of input raw data was performed using a sliding windows (size $2.56\,s$, $50\%$ overlap) [40]. For each sliding window, 128 samples were supplied from the six channels (three axes signal from accelerometer and gyroscope). A stochastic gradient descent optimiser was employed. Training was performed using dropout layers (rate set at 0.5), following each convolutional layer. The classifier was trained in 1000 epochs. Model weights for minimal loss on the validation set were kept. For multiple training iterations, the network was observed to reach minimum loss on the validation set after 400-500 epochs.

### 6.2 Sensor Dataset

The analysis was carried out on the public available dataset published by Anguita et al. [41]. The dataset includes recordings of 30 healthy volunteers, aged between 19 and 40 years, who performed six activities while wearing a smartphone on the waist. 3-axial linear acceleration and 3-axial angular velocity were recorded from embedded accelerometer and gyroscope at a constant rate of 50 Hz. At fixed-width sliding windows of 2.56 sec and 50% overlap, 10299 instances can be used for classification learning. The dataset was published on the UCI machine learning repository in 2012 and it has been used as classification benchmark in various papers, e.g., [42], [43].

### 6.3 Validation Strategy

Five classes were considered: stationary (i.e., standing and sitting), transition, walking, walking upstairs and walking downstairs. Evaluation was performed with the data of 18 participants for training, three for validation, and nine for testing. The CA-SNS hyperparameters were chosen manually on the validation dataset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



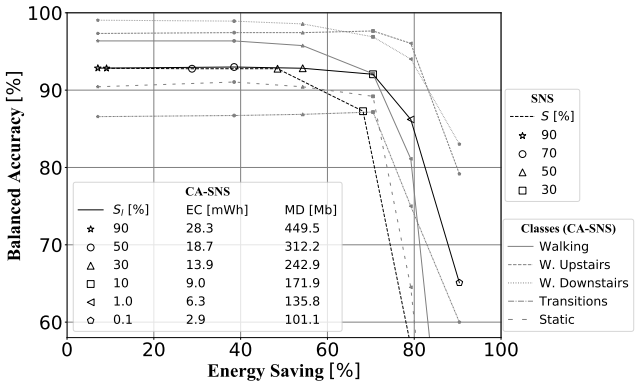


Fig. 7: Case study of Sec. 6, showing physical activity classification results derived on the validation set. Parameters: $N = 128$, $G = 6400$, $S_h = 1$, $\theta_l^{acc} = 1.9$, $\theta_l^{gyr} = 0.20$, $\theta_h^{acc} = 4$, $\theta_h^{gyr} = 4$, $J = 16$, $F_M = 11$, $F_y^{acc} = 0.1$, $F_y^{gyr} = 0.03$, $\Omega = 25$.

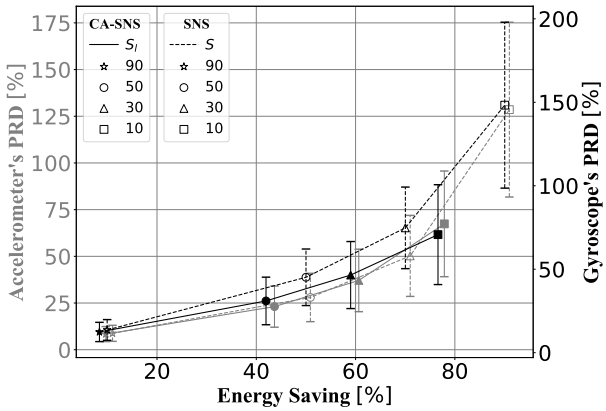


Fig. 8: Case study of Sec. 6. Parameters: $N = 128$, $G = 6400$, $S_h = 1$, $\theta_l^{acc} = 1.9$, $\theta_l^{gyr} = 0.2$, $\theta_h^{acc} = 4$, $\theta_h^{gyr} = 4$, $J = 16$, $F_M = 11$, $F_y^{acc} = 0.1$, $F_y^{gyr} = 0.03$, $\Omega = 25$. Results derived on the validation set.

### 6.4 Results

The three reconstruction algorithms showed similar performance, as in the case study of Sec. 5, thus we focused on results based on OMP. Fig. 7 depicts the tradeoff between classification performance and energy saving. In CA-SNS, the accuracy was kept over 90% while achieving an energy saving up to $75\%$. In SNS, the accuracy dropped above an energy saving of 50%. At an energy saving $\sim 75\%$, the difference between CA-SNS and SNS accuracy was $\sim 10\%$. Furthermore, Fig. 7 depicts the tradeoff between classification performance and energy saving for individual classes. No significant recognition performance degradation up to an energy saving of $\sim 75\%$ was observed for all classes. Fig. 8 compares PRD and energy saving tradeoff for the reconstructed accelerometer and gyroscope signals. The CA-SNS achieved a lower PRD than SNS in the gyroscope signal reconstruction. No significant difference was found between CA-SNS and SNS PRD in the accelerometer signal reconstruction. Tab. 5 shows the results summary derived on test dataset. At similar retrieval performance, CA-SNS saved approx. $22\%$ in energy and memory compared to SNS.

| | **Full Sampling** mean (std) | **SNS** mean (std) | **CA-SNS** mean (std) | **Stat. Sig. Test** (Significance Level = 5%) |
|---|---|---|---|---|
| Balanced Acc. [%] | 92.4 (1.7) | 92.7 (3.8) | 92.1 (6.3) | Not significant |
| PRD (Acc.) [%] | - | 50.2 (21.6) | 67.3 (28.2) | Not significant |
| PRD (Gyro.) [%] | - | 74.7 (24.4) | 70.1 (29.9) | Not significant |
| Energy Saving [%] | - | 48.5 (0.4) | 70.5 (1.2) | $10^{-7}$ |
| EC [mWh] | 30.4 (3.1) | 12.9 (0.8) | 10.1 (1.6) | $10^{-6}$ |
| MD [MB] | 483.6 (3.4) | 215.7 (0.8) | 168.3 (1.8) | $10^{-7}$ |

TABLE 5: Case study of Sec. 6, summarising physical activity classification results derived on testing data. Results were averaged across ten evaluations, reporting mean and standard deviation (std). Parameters: $N = 128$, $G = 6400$, $S_h = 1$, $S_l = 10$, $S = 30$, $\theta_l^{acc} = 1.9$, $\theta_l^{gyr} = 0.2$, $\theta_h^{acc} = 4$, $\theta_h^{gyr} = 4$, $J = 16$, $F_M = 11$, $F_y^{acc} = 0.1$, $F_y^{gyr} = 0.03$, $\Omega = 25$.

## 7 Case Study: Heartbeat Classification Using Single-Lead ECG

A 12-lead ECG is composed by a three bipolar limb leads (I, II, and III), three unipolar limb leads (AVR, AVL, and AVF), and six unipolar chest leads (V1, V2, V3, V4, V5, V6). Morphological anomalies in the QRS-complex relate to type of arrhythmias, which can be detected by ECG heartbeat classification.

### 7.1 Pattern Recognition Method

QRS-complex detection is a challenging task due to artifacts in the signal as muscle movements, power line interference, or motion. We employ a QRS detection algorithm inspired Pan and Tompkins [44]. An ensemble of Support Vector Machines (SVMs) classified normal beats and beats affected by different types of arrhythmia. The classification method was introduced by Mondéjar-Guerra et al. [45]. Separate SVMs were trained with feature based on R-R intervals, wavelets, higher order statistics, local binary patterns, and several amplitude values. A combinatorial product rule was used to merge the predictions. The feature and ensemble details can be found in [45].

### 7.2 Sensor Dataset

Data were collected form 47 patient (15 females, 17 males), aged between 18 and 80, undergoing tests for coronary artery disease. Most of the patients had ventricular ectopic beats. The patient selection process preferred subjects with ischemia, coronary disease, conduction abnormalities, and arrhythmias. The dataset was composed of 44 ECG records of $\sim 30\,m$ measured at a uniform

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



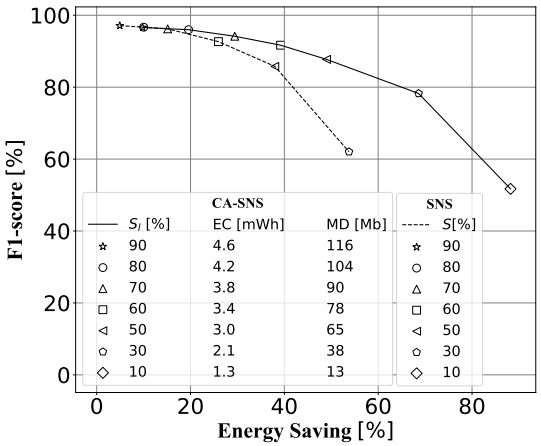


Fig. 9: Case study of Sec. 7, showing heartbeat detection results derived on the validation set. F1-score vs. energy saved for CA-SNS and for SNS. Parameters: $N = 120$, $G = 6400$, $S_h = 1$, $\theta_l = 500$, $\theta_h = 8000$, $J = 12$, $F_M = 6$, $F_y = 30$, $\Omega = 25$.

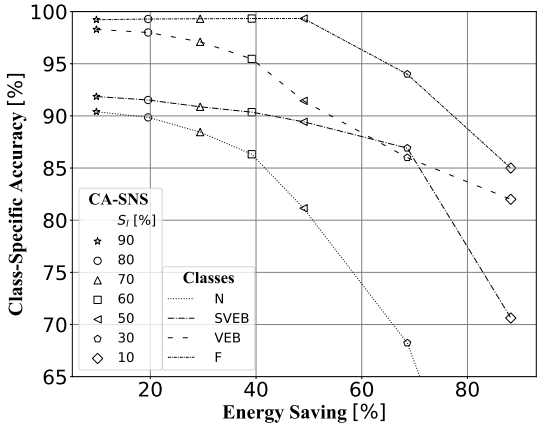


Fig. 10: Case study of Sec. 7, showing arrhythmias anomaly classification results derived on the validation set. Parameters: $N = 120$, $G = 6400$, $S_h = 1$, $\theta_l = 500$, $\theta_h = 8000$, $J = 12$, $F_M = 6$, $F_y = 30$.

sampling rate of 360 Hz. The modified lead II (MLII) was considered, in line with the most common practice in cardiac arrhythmia analysis.

### 7.3 Validation Strategy

We evaluated heartbeat detection and arrhythmias anomaly classification performance against annotated heartbeat events. Four classes were considered: Normal beat (N), Supraventricular beat (SVEB), Ventricular beat (VEB), Fusion beat (F). Evaluation was performed with the data of 22 ECG records for training, 20 for validation, and two for testing. The CA-SNS hyperparameters were chosen manually on the validation set. The hyperparameter optimisation of QRS-complex detection was performed using grid search approach. The SVM hyperparameters were derived from [45].

### 7.4 Results

The three reconstruction algorithms showed similar performance, as in the case studies of Sec. 5 and 6, thus we focused on results based on OMP. Fig. 9 shows the tradeoff between heartbeat detection performance and energy saving. With CA-SNS minimum sampling probability $S_l = 60\%$ and SNS sampling probability $S = 60\%$, a difference in energy saving of $\sim 15\%$ was estimated, with a similar F1-score of $\sim 92\%$. Fig. 10 focuses on CA-SNS to illustrate the class-dependent performance for arrhythmia classification. Tuning the minimum sampling probability $S_l$ had different effects on the individual classes. Classification accuracy of F and SVEB classes were maintained at $\sim 98\%$ and $\sim 90\%$ respectively, while achieving an energy saving $\sim 50\%$. In contrast, classification accuracy of N and VEB classes dropped as $S_l$ was reduced below $70\%$.

Fig. 11 compares PRD and energy saving tradeoff for the reconstructed ECG signals. Overall, CA-SNS achieved a lower PRD than SNS. The higher the sensitivity $S_l$, the higher the difference between signal reconstruction PRDs of CA-SNS and SNS.

Tab. 6 shows the results summary derived on test dataset. At similar retrieval performance, CA-SNS saved approx. $13.3\%$ in energy and memory compared to SNS.

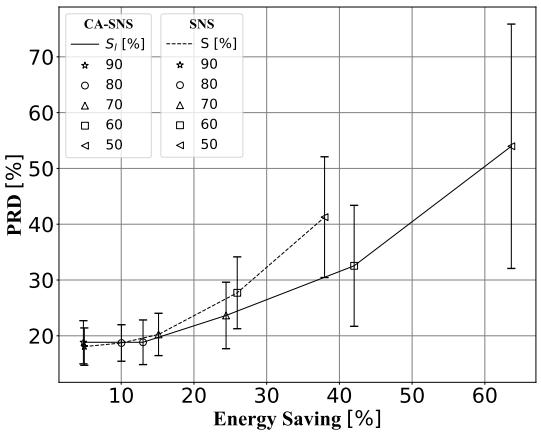


Fig. 11: Case study of Sec. 7. Parameters: $N = 120$, $G = 6400$, $S_h = 1$, $\theta_l = 500$, $\theta_h = 8000$, $J = 12$, $F_M = 6$, $F_y = 30$, $\Omega = 25$. Results derived on the validation set.

## 8 Discussion

Wearable sensing systems are characterised by their limited computational power, communication bandwidth, memory capacity, and energy budget. Reducing the samples retrieved from the sensors can be beneficial for all system resources.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3077731, IEEE Transactions on Mobile Computing



| | **Full Sampling** mean (std) | **SNS** mean (std) | **CA-SNS** mean (std) | **Stat. Sig. Test** (Significance Level = 5%) |
|---|---|---|---|---|
| F1-Score [%] | 98.3 (2.1) | 92.1 (5.3) | 91.6 (7.2) | Not significant |
| Balanced Acc. [%] | - | 93.4 (2.3) | 92.8 (5.2) | Not significant |
| PRD [%] | - | 29.5 (6.4) | 35.4 (10.8) | Not significant |
| Energy Saving [%] | - | 28.9 (0.5) | 42.2 (6.7) | $10^{-4}$ |
| EC [mWh] | 5.2 (1.7) | 3.81 (0.4) | 3.3 (0.9) | $10^{-3}$ |
| MD [MB] | 130.4 (1.9) | 91.3 (0.7) | 79.5 (1.4) | $10^{-4}$ |

TABLE 6: Case study of Sec. 7, summarising heartbeat detection and arrhythmias anomaly classification results derived on testing data. Results were averaged across ten evaluations, reporting mean and standard deviation (std). Parameters: $N = 120$, $G = 6400$, $S_h = 1$, $S_l = 60$, $S = 60$, $\theta_l = 500$, $\theta_h = 8000$, $J = 12$, $F_M = 6$, $F_y = 30$, $\Omega = 25$.

Our approach is driven by the dynamic variation of signal information rate in the sensor data of most wearable systems. While the CS concept is well-known, methods to estimate signal context, design response models, as well as integrate dynamic sensing matrix compression, have not been explored before. The significant energy saving over standard SNS across three case studies with different sensors and signal properties confirm the potential of our CA-SNS.

In the case study of Sec. 5, we used a three-way cross-validation to obtain the performance results. For the subsequent case studies (Sec. 6 and 7), we followed the validation procedures used for the respective datasets in earlier publications. Maintaining the same validation procedures enabled us to confirm that the recognition performances remain comparable to earlier publications using the datasets.

CA-SNS yields a better tradeoff between energy saving and recognition performance than SNS. The performance gap however varies with the signal characteristics and hardware. Applications related to physical activity (case study of Sec. V using EMG and case study of Sec. VI using inertial sensors) benefitted from the rare signal information rate increases, thus yielding performance increases of $17\%$ and $22\%$. In contrast, for periodic signals (case study of Sec. VII using ECG) and application-specific hardware, the performance improvement was only $13\%$. We observed that the reactive response model of CA-SNS tends to miss samples from the beginning of the QRS-complex, which led to recognition performance degradation when reducing response model sensitivity $S_l$. Clearly, the context estimate, set as the same basic function across all case studies, limits the CA-SNS benefits for periodic signals. More generally, adaptability to context variability is a key for acceptable reconstruction error in CA-SNS, which affects the downstream data analysis, e.g., pattern recognition algorithms. In the resource-constrained setting of mobile and wearables devices, signal approximation must be computationally inexpensive and therefore could only involve low-order signal moments. A naive first choice to represent a future signal sample is its previous value, i.e., local expectation about the signal, and use signal variance, i.e. the signal's first order moment, as a proxy of information entropy. Intuitively, when a signal varies between samples, information entropy is larger than for situations, where signal variance is minimal, i.e. new signal samples are similar to past ones. In our analysis, we adopted a well-understood context estimate $f_C(y_t) = var(y_t)$, which is applicable to various physiological signals. We applied a one-pass streaming algorithm to calculate $var()$ online, without unbounded buffering in $\mathcal{O}(1)$ time and storage. However, signal variance is susceptible to signal artefacts and noise and thus may unnecessarily increase sampling probability. Nevertheless, our results indicate that signal variance as context estimate yields robust results across all signal types and case studies considered here. Future research may further investigate the context estimate function and response model design to leverage signal-specific features and hardware configurations.

| **Case study** | **Runtime** [h] | | |
|---|---|---|---|
| | Full Sampling | SNS | CA-SNS |
| Spotting chewing pattern using EMG-monitoring eyeglasses (Sec. 5) | 2.1 (256 Hz) | 9.6 | 11.6 |
| Human activity recognition using waist-worn inertial sensors (Sec. 6) | 32.6 (50 Hz) | 80.1 | 102.7 |
| Heartbeat classification using single-lead ECG (Sec. 7) | 177.8 (360 Hz) | 242.7 | 280.3 |

TABLE 7: Estimated runtime for all the case studies considering a battery capacity of 925 mWh.

The behaviour of CA-SNS depends on the hyperparameter set. The time period length $N$, the size $G$ of the overcomplete dictionary, the target sparsity $\Omega$, and the information rate classifier parameters $F_M$ and $F_y$, were defined experimentally through observations. The response model hyperparameters $S_h$, $\theta_h$, and $\theta_l$, were found via grid-search. For simplicity, we adopted an offline dictionary learning strategy and trained overcomplete dictionaries using the K-SVD algorithm. Nevertheless, online optimisation algorithms [46] may be applied to train basis sets continuously with potential benefits in free-living, e.g., adaptation to concept drift. Appendix D reports further details on the hyperparameter selection.

In our simulation, we assumed a 32-bit sample resolution for all case studies. Nevertheless, the relation of CA-SNS and SNS would remain the same even if other sample resolutions were used. We made efforts to accurately model the computational complexity. Nonetheless, the costs of generating random numbers for the sampling

was not included, as the total random number count was the same for SNS and CA-SNS and thus did not have an effect on the relative performance analysis. In our simulations, a random number was generated for every sample in a dataset and then a CA-SNS or SNS-specific Bernoulli outcome derived. Similarly, the hardware properties have a minor effect on the performance relation. Even if other hardware was considered in the case studies, the advantage of the CA-SNS approach would persist.

In this work, we simulated a continuous data stream radio transmission through Chipcon transceiver CC2420. However, packet-oriented radios (like CC2420) could increase efficiency through burst transfers. CA-SNS supports burst transfers as the signal reconstruction works per time period $N$. Thus, burst could further reduce radio energy $EC^r$ compared to continuous transmission. The approach presented in this work addresses low resource devices that provide sensor data to an integration node for processing. Nevertheless, we believe that the CA-SNS concept can be adapted to systems that perform local processing, either for resource-intensive sensors (energy / bandwidth) or by adapting pattern recognition algorithms to work with the sparse signal representation.

The idea of CA-SNS is to select non-equidistant samples from sensors and let the devices sleep in the remaining time. Our approach corresponds to an analogue CS (aCS) encoder, i.e., compress signal at the sampling stage. While aCS can be realised in most technical systems including those of our case studies, CA-SNS can be applied for digital CS (dCS) too. dCS compresses after all samples of the original signal have been obtained. In a dCS setup, CA-SNS can still save radio energy by transmitting a compressed signal.

A reduction of current required from the battery can be used either to reduce battery size in the system design phase or extend runtime during deployment. Tab. 7 summarises the estimated runtimes across case studies depending on the sampling strategy.

## 9 Conclusion

This work proposed a context-adaptive sample acquisition strategy based on the sub-Nyquist framework for sampling to minimise energy requirements in wireless embedded sensing systems. We consider that ultra-low power mobile and wearable sensing systems are particularly exposed to contextual variations and thus can benefit from our approach. Across three case typical studies, we analyse the potential of CA-SNS in detail and compare to full sampling and standard SNS. Our results clearly demonstrate that CA-SNS yields profound energy saving between 13% and 22% over SNS, while pattern recognition and signal reconstruction performances are maintained. With our results, longer runtimes or reduced battery and system sizes can be realised.

## Acknowledgement

This work has been partially funded by the EU H2020 MSCA ITN ACROSSING project (GA no. 616757).

## References

[1] O. Amft, "How wearable computing is shaping digital health," *IEEE Pervasive Computing*, vol. 17, no. 1, pp. 92–98, 2018.

[2] E. J. Candès *et al.*, "Compressive sampling," in *Proceedings of the international congress of mathematicians*, vol. 3, pp. 1433–1452, Madrid, Spain, 2006.

[3] R. E. Carrillo, A. B. Ramirez, G. R. Arce, K. E. Barner, and B. M. Sadler, "Robust compressive sensing of sparse signals: a review," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 108, 2016.

[4] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, 2012.

[5] X. Chen, E. A. Sobhy, Z. Yu, S. Hoyos, J. Silva-Martinez, S. Palermo, and B. M. Sadler, "A sub-nyquist rate compressive sensing data acquisition front-end," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 542–551, 2012.

[6] M. Mishali and Y. C. Eldar, "Sub-nyquist sampling," *IEEE Signal Processing Magazine*, vol. 28, no. 6, pp. 98–124, 2011.

[7] D. Brunelli and C. Caione, "Sparse recovery optimization in wireless sensor networks with a sub-nyquist sampling rate," *Sensors*, vol. 15, no. 7, pp. 16654–16673, 2015.

[8] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.

[9] G. Pastor, I. Mora-Jiménez, R. Jäntti, and A. J. Caamano, "Mathematics of sparsity and entropy: Axioms core functions and sparse recovery," *arXiv preprint, https://arxiv.org/abs/1501.05126v3*, 2015.

[10] P. Charalampidis, A. G. Fragkiadakis, and E. Z. Tragos, "Rate-adaptive compressive sensing for iot applications," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2015.

[11] Z. Liu, W. Yu, C. Chen, B. Yang, and X. Guan, "Adaptive compression ratio estimation for categorified sparsity in real-time ecg monitoring system," in *2016 35th Chinese Control Conference (CCC)*, pp. 9439–9444, IEEE, 2016.

[12] Y. He, W. Yu, C. Chen, Y. Wang, and X. Guan, "Adaptive ecg compression scheme with prior knowledge support based on compressive sensing," in *2015 International Conference on Wireless Communications & Signal Processing (WCSP)*, pp. 1–5, IEEE, 2015.

[13] V. Behravan, N. E. Glover, R. Farry, P. Y. Chiang, and M. Shoaib, "Rate-adaptive compressed-sensing and sparsity variance of biomedical signals," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–6, IEEE, 2015.

[14] S. Traore, B. Aziz, D. Le Guennec, and Y. Louet, "Adaptive non-uniform sampling of sparse signals for green cognitive radio," *Computers & Electrical Engineering*, vol. 52, pp. 253–265, 2016.

[15] S. Engelberg, *Digital signal processing: an experimental approach*. Springer Science & Business Media, 2008.

[16] R. Fallahzadeh, J. P. Ortiz, and H. Ghasemzadeh, "Adaptive compressed sensing at the fingertip of internet-of-things sensors: An ultra-low power activity recognition," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*, pp. 996–1001, IEEE, 2017.

[17] B. Welford, "Note on a method for calculating corrected sums of squares and products," *Technometrics*, vol. 4, no. 3, pp. 419–420, 1962.

[18] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[19] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, 2007.

[20] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[21] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of selected topics in signal processing*, vol. 1, no. 4, pp. 586–597, 2007.

[22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[23] L. F. Polania and K. E. Barner, "Multi-scale dictionary learning for compressive sensing ecg," in *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, pp. 36–41, IEEE, 2013.

[24] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[25] M. Er, "Lexicographic ordering, ranking and unranking of combinations," *International journal of computer mathematics*, vol. 17, no. 3-4, pp. 277–283, 1985.

[26] V. Parque and T. Miyashita, "Unranking combinations using gradient-based optimization," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 579–586, IEEE, 2018.

[27] M. Er, "Lexicographic enumeration, ranking and unranking of permutations of r out of n objects," *International Journal of Computer Mathematics*, vol. 23, no. 1, pp. 1–7, 1987.

[28] M. Buschhoff, C. Günter, and O. Spinczyk, "A unified approach for online and offline estimation of sensor platform energy consumption," in *2012 8th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 1154–1158, IEEE, 2012.

[29] ARM limited, https://developer.arm.com/documentation/100165/[illegible], *Cortex-M3 Technical Reference Manual*, 2006.

[30] CamNtech, https://www.camntech.com, *Actiwave, CamNtech*.

[31] InvenSense Inc., https://invensense.tdk.com/wp-content/uploads/2015/02/PS-MPU-9250A-01-v1.1.pdf, *MPU-9250 Product Specification*, 2016.

[32] Texas Instruments, https://www.ti.com/tool/TIPD116, *Ultra Low Power, 18 bit Precision ECG Data Acquisition System*, 2013.

[33] Chipcon Products from Texas Instruments, https://www.ti.com/product/CC2420, *2.4 GHz IEEE 802.15.4 / ZigBee-ready RF Transceiver*, 2013.

[34] A. Prayati, C. Antonopoulos, T. Stoyanova, C. Koulamas, and G. Papadopoulos, "A modeling approach on the telosb wsn platform power consumption," *Journal of Systems and Software*, vol. 83, no. 8, pp. 1355–1363, 2010.

[35] C. Nadeau and Y. Bengio, "Inference for the generalization error," in *Advances in neural information processing systems*, pp. 307–313, 2000.

[36] R. Zhang, S. Bernhart, and O. Amft, "Diet eyeglasses: Recognising food chewing using emg and smart eyeglasses," in *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 7–12, IEEE, 2016.

[37] R. Zhang and O. Amft, "Monitoring chewing and eating in free-living using smart eyeglasses," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 23–32, 2017.

[38] R. Zhang and O. Amft, "Free-living eating event spotting using emg-monitoring eyeglasses," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 128–132, IEEE, 2018.

[39] F. Cruciani, A. Vafeiadis, C. Nugent, I. Cleland, P. McCullagh, K. Votis, D. Giakoumis, D. Tzovaras, L. Chen, and R. Hamzaoui, "Comparing cnn and human crafted features for human activity recognition," in *2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing UIC*, Aug 2019.

[40] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.

[41] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *International workshop on ambient assisted living*, pp. 216–223, Springer, 2012.

[42] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert systems with applications*, vol. 59, pp. 235–244, 2016.

[43] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4566–4578, 2016.

[44] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Trans. Biomed. Eng*, vol. 32, no. 3, pp. 230–236, 1985.

[45] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ecgs via ensemble of classifiers," *Biomedical Signal Processing and Control*, vol. 47, pp. 41–48, 2019.

[46] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: algorithms and applications," *IEEE access*, vol. 3, pp. 490–530, 2015.


**Giovanni Schiboni** received a B.Sc. and a M.Sc. in biomedical engineering at Universita' Campus Bio-Medico, Rome (Italy), between 2004 and 2011. In 2015, he received a M.Res. in Advanced Engineering Robotics at Bristol Robotics Laboratory, University of West England. As former ESR Marie Curie fellow, part of the ACROSSING project (http://www.acrossing-itn.eu/), he is pursuing his PhD degree at the Chair of Digital Health, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen (Germany). His research project focuses on energy-efficiency of wearable sensing technology applied to automatic dietary monitoring. Currently, Giovanni is Chief Data Scientist at the swiss med-tech startup Inanna Fertility (Zürich), improving assisted reproductive technology outcomes using machine learning over medical big data.

**Celia Martin Vicario** received a B.Sc. in biomedical engineering at Technical University of Madrid in 2017 and she completed a M.Sc. in medical engineering in Friedrich-Alexander Universität of Erlangen-Nürnberg in 2020. In 2020 she joined the Pattern Recognition Lab at the Friedrich-Alexander Universität of Erlangen-Nürnberg as a PhD candidate. She is currently involved in scientific research in the fields of image processing and deep learning.

**Juan Carlos Suarez** received his BSc. degree in Medical Engineering from FAU Erlangen-Nürnberg in 2019. Currently, he is pursuing his MSc. degree in Medical Engineering at FAU Erlangen-Nürnberg. He works as a research assistant at the Chair of Digital Health. His research interests are in the field of machine learning and pattern recognition applied to biomedical signals from wearable devices.

**Federico Cruciani** received his BSc. Eng. and MSc. Eng. in Computer Science Engineering at University of Florence, Italy. From 2006 to 2016 he worked as software project manager at I+ and Orthokey focusing on Computer Assisted Surgery and eHealth, working on optical and inertial sensors applied to kinematic analysis and rehabilitation. In 2016, he joined Ulster University as Marie Curie fellow in Computer Science, where he received the PhD degree in 2020. He is currently working as Research Fellow at Ulster University.

**Oliver Amft** is the founding director of the Chair of Digital Health at the Friedrich Alexander University Erlangen-Nürnberg (FAU). He received the M.Sc. from Chemnitz Technical University in 1999 and the Ph.D. from ETH Zurich in 2008, both in Electrical Engineering and Information Technology. In 2011, he obtained the university teaching qualification from TU Eindhoven. Until 2004, he was a R&D project manager with ABB, Inc. Oliver Amft was an assistant professor at TU Eindhoven between 2009 and 2013 and full professor directing the Chair of Sensor Technology at University of Passau since 2014. Since 2017, Oliver is a full professor at FAU. Oliver is interested in context recognition, wearable sensor technology, and design of biomedical systems and has co-authored over 200 publications in this area.